最小距离分类器(minimun-distance classifier)

原理

利用了 *Nearest-neighbour decision rule* 的思想 : 假定现在有n类 $c_i \in c1, c2, ..., c_n$

给每一个类可以选出一个 **类代表** (可以是每一个类的中位数(the mean of class) etc.) , 对应的记为 $p_i \in p1, ... p_n$

这里的 P_i 维度必须和**输入数据的属性维度一样**, 既可以看作可输入数据一样的在**坐标轴上的一个个的点**, 这些点就分别代表了他们各自的类

现在Classifier的思想就是 : **输入一个数据 *X* , 分别计算这个数据和各个点 P_i 的距离[1], 取距离最近的那一类归位这个点的类**

[1] : 这个距离可以为 **欧式距离, 曼哈顿距离, etc.**, 但是一般在坐标轴上就是欧式距离

实现公式

假设输入点为 *X* = [x1, x2, ... , x$_n$] , 点P_i为 P = [p1, p2, ..., p$_n$]

那么这两点的距离可以记为 : Dis(X, P$_i$) = |X - P$_i$ |^2 = X^TX - 2X^TP_i + $P_i^TP_i$

因此我们只需要一次计算 X 和 各个 P$_i$ 之间的取最小的距离即可

Min$_i$ Dis(X, P$_i$) i $\in$ 1, 2, ... n (number of classes), 其实观察之后可以发现每个距离的第一项 X^TX 等相等, 则可以不用计算, 式子变为 :

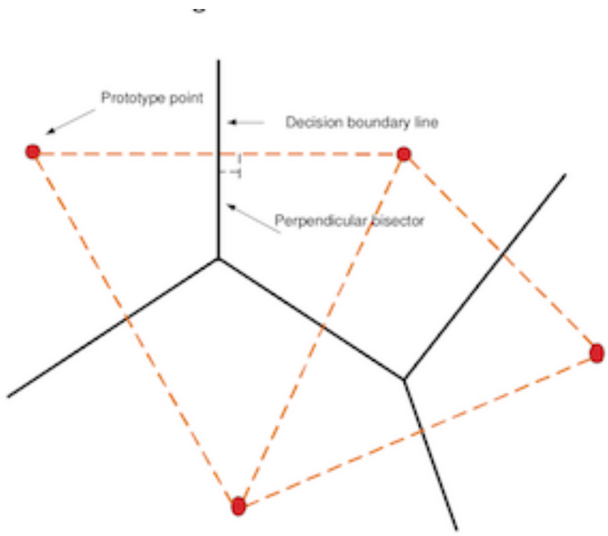
The class assigned to is :

w_i = min$_i$(- 2X^TP_i + $P_i^TP_i$) = max$_i$(X^TP_i - $\frac{1}{2}P_i^TP_i$)

结合到我们线性分类器的公式中 :

g$_i$(x) = w$_i^t X$ + w_{i0} (w_i = p_i , w$_{i0}$ = -$\frac{1}{2}|p_i|^2$)

图示



- 可以看到经过最小距离形成的决策界就是两点连线之间的垂线
- 且这个线性分类决策区域是 **always convex** 的